CHAPTER V

Figure 1: Distributions of the best features which were part of the TD set mentioned in Chapter IV. This dataset included physics based feature transformations not unlike common CTMs. These proposed features, however, are much simpler to encode and include in ML/AI ensembles. Each of the variables used are representative of chemical reactions and O3 trends mentioned in Chapters II and III.

RESULTS

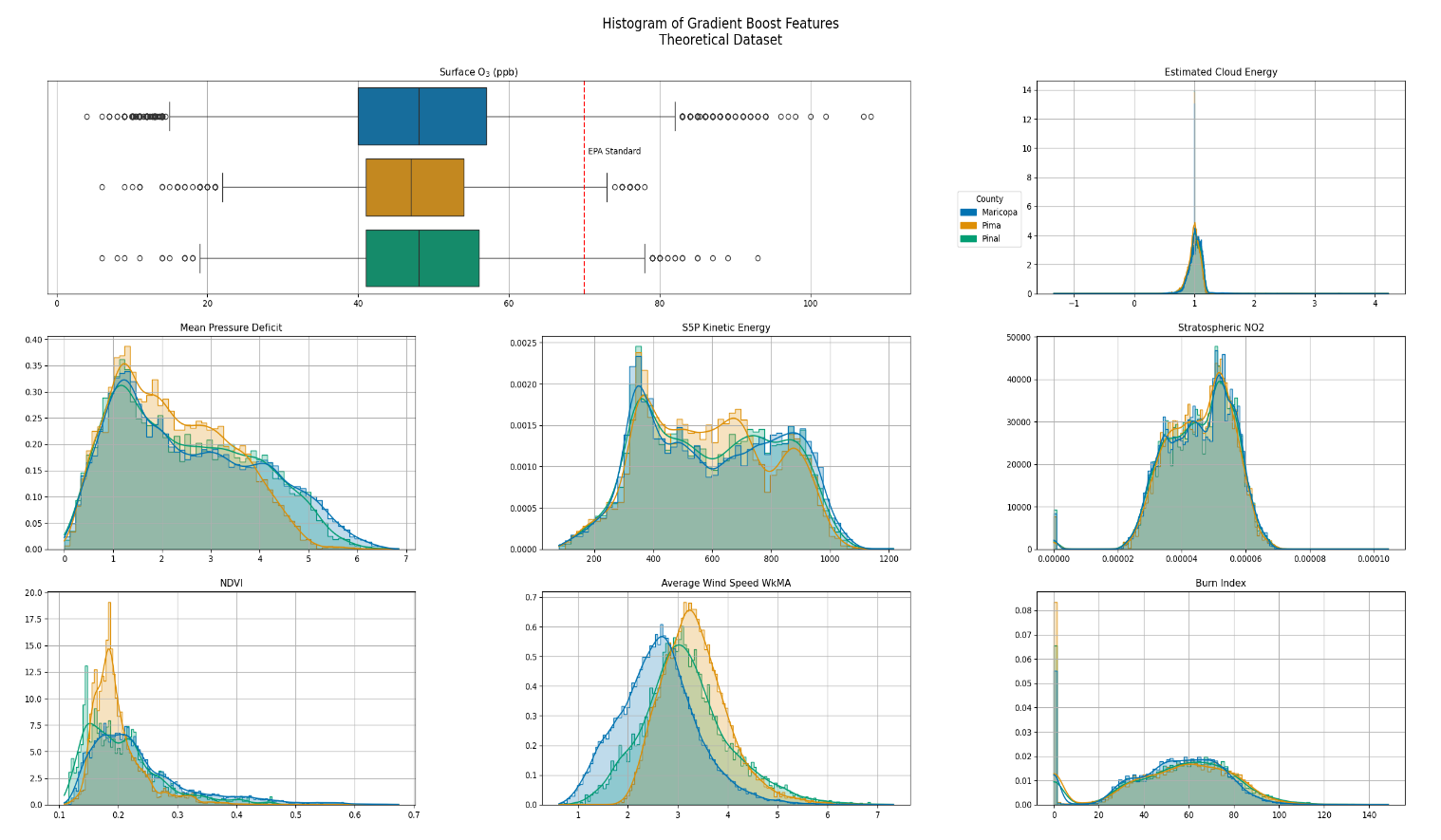
**Variable codes for this chapter are given at VIII.1.3 in the appendix. While there were 44 features extracted from the sources in Chapter IV, only the top 24 best correlative features were used at once for each ensemble. Seasonal dummy variables were included with Fall having the least correlative power among them. This was consistent with trends mentioned in Chapter II e.g. (M. Li et al. 2021; Seroji 2016; Zhao, Zheng, and Li 2018). As such, Fall was omitted from the models, establishing a baseline for the overall trend. Monthly dummy variables initially created were too recursive of a feature. Many of the months outside of winter and summer seasons were not as well correlated. Many of the best ensemble and RK tuning parameters are found to be representative of the diligent preprocessing work and pseudo-CTM based feature transformation, with TD having the best outcomes. G.O.A.T.24 offered suitable results and could have been utilized as well, however the results of each model were more consistent with TD overall. HD and MD yielded extremely similar errors, indicating the possibility of a readily available historical dataset from 1980-present day.

Figure V.1

Final Gradient Boost features which yielded the least error are depicted across each county in PHOTUC. On average, Maricopa sees more values higher than the EPA standard set at 70 ppb. In addition, to having lower wind speeds Maricopa has slightly higher NDVI values while Pima tends to have significantly lower values around 0.2. Maricopa and Pinal which house Phoenix and Tucson respectively might have healthier vegetation on average due to a larger number of resources spent towards maintaining urban greenspace.

The SMaRK method yielded the least error across each of the datasets mentioned in Chapter VI.4.3. The full results can be seen in Figure 10, with SMaRK based gradient boosting yielding the best results for this project. Each ensemble, when tuned properly, better modeled the known trend with the incorporation of spatial uncertainty as opposed to reporting it in the error calculations. In addition, the distribution of resulting errors is the same, but overall accuracy is significantly reduced, indicating strong positive reinforcement from the spatial krige across all models. Ensembles like gradient boost, random forest, and perceptron yielded similar RMSE to studies utilizing sole ML ensembles such as (K. Anand et al. 2025; C T et al. 2020; Chen et al. 2023; Huang et al. 2025; Kleinert, Leufen, and Schultz 2021; Ko, Cho, and Rao 2022; Q. Li et al. 2024; Mu et al. 2023; Nelson et al. 2023).

RK enhancements to these already low predictions further reduced error by adjusting for geospatial uncertainty given the trends of selected features over the AOI. Physics and CTM based features were among the highest predictive features and should be further investigated. These are further compared in Chapter VI and VIII, which are dedicated to reasoning and analysis of modern efforts to model surface O3 concentrations. Figure 9 depicts the Pearson Correlation Matrix which represents the correlation of 31 of the 47 features mentioned in chapter IV. While every feature gathered was correlated to surface O3 production before the training process was conducted, only the more prominent features showed deviations in SHAP plots depicted in XXXXX.

The full set of training features can be seen in the appendix; a sample of them with features in the final prediction and the distribution of the average surface O3 in PHOTUC are noted in Figure V.1. They were used mainly due to this excerpt from Figure VIII.2.10.; depicting datasets and their best output from SMaRK predictions:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Feature Count** | **Variable Code** | **MAPE** | **MAE (ppb)** | **RMSE**  **(ppb)** | **SMaRK Base** |
| Historical  (H.D.) | 13 | V1, V2, V3, V5, V6, V10, V11, V13, V16, V24, V25, V26 | 2.67% | 1.19 | 0.026 | Random Forest |
| Modern  (M.D.) | 13 | V1, V2, V3, V4, V5, V6, V7, V8, V9, V24, V25, V26 | 5.91% | 2.62 | 0.12 | Extreme Gradient Boost |
| Theory  (T.D.) | 12 | V4, V5, V6, V9, V13, V18, V22, V23, V24, V25, V26 | 1.39% | 0.59 | 0.013 | Gradient Boost |
| G.O.A.T.24 | 24 | All features except V22 and V23 | 5.71% | 2.55 | 0.12 | Extreme Gradient Boost |

Table V.1

Variable codes are found under VIII.3. GOAT24 variables did not include NVDI and estimated LN cloud energy due to their low correlations under 0.2. The absolute correlation of all features was taken and used to create each dataset representative of available features in a certain timeframe.

TD had the best model outputs and lowest effective error across all metrics. SMaRK enhanced gradient boost was utilized for PHOTUC due to its consistency across model outputs, yielding errors form around 0.59 ppb ± 0.013 ppb to 2.78 ppb ± 0.14 ppb. All statistical models yielded around 10% error overall, ranging from 1.47% to 14.18% error with variations of proper and improper tuning of the searching algorithm. The ML/AI ensembles alone predicted surface O3 concentrations in PHOTUC with an MAE of around 0.61 ppb to 6.02 ppb, like studies using them with CTM based features e.g. (Chen et al. 2023; Hu et al. 2022; Ko, Cho, and Rao 2022). RK enhancements incorporating geospatial uncertainty into the ML/AI models yielded between 1.39%-6.36% error, with MAE values of 0.59 ppb to 2.82 ppb. SMaRK-modeled surface O3 benefited all model ensembles over the course of study.

Each ensemble: including early linear regression and weighting methods; ran hundreds of times over the course of this project to test reproducibility of the script and computational efficiency. Machines of varying computational processing power evaluated practicality and usability of SMaRK. Three main setups were tested: an Apple A1 MacBook Air 2020, a Windows Desktop with a Ryzen 5700XT Processor, and a Citrix based Cloud Computing Desktop provided by the Institute of Behavioral Sciences (IBS) at CU, Boulder. Many external functions in a library developed with python 3.12.x (Rossum and Drake 2010) upload to GitHub can be easily imported on user machines and implemented for various AOIs.

V.1 Proposed Feature Transformations, ML/AI Ensembles, and RK

The unique functions allowed for rapid creation of several predictive rasters from January 1st, 2019, to December 16th, 2024, over the AOI. The 16th represents the stopping point of the raster interpolation strategy mentioned in Chapter III.2.2. Pre-processing methods for this thesis are reproducible for data from 1980-present day. Some of the raster pre-processing steps, integration, and display functions are currently in development as Python continues development, shifting into well upgraded functions with more, if not better, user accessibility, e.g. Rasterio (Gillies and others 2013) based in GDAL (Rouault et al. 2025) for raster sampling and extraction. The distributions of training features, interpolation strategies, model parameters, kriging enhancements, and case-study statistics are expanded upon in detail. Further review of the final predictive raster displays, model outputs, and relations to Chapter II are in Chapter VI. Despite the These overall errors in predictions, with lack-luster incorporation of training values in each geo-field were lower than that of similar CTM and emission transport representations of surface O3 concentrations e.g.(Chattopadhyay, Midya, and Chattopadhyay 2019; Pan, Harrou, and Sun 2023; Zhang et al. 2022). where the R2, RMSE, and MAE of proposed complex ensembles were around 0.9, 15ppb, and 9 ppb respectively.

Data sets like G.O.A.T 25, which included more features to increase the amount of varying dependent variables, still produced usable models with less error than the statistical model alone, albeit exponentially increasing computation times. The minimal change in error regarding HD and ND models shows the creation of a large, historical database which covers the USNA is possible utilizing SMaRK methods and feature transformations. The selected time frame contained 30 out of 38 unique monitors distributed across the PHOTUC region depicted in Figure X. Monitors with about 80% of missing data, namely due to age or monitor malfunctions, were incapable of interpolation and imputation. After interpolating available monitors in the PHOTUC region, 5,225 dates per 30 monitors for a total of 65,250 data-points were utilized on model training. The two monitor locations omitted did have more data available before 2018 and could be used in the historical model making process.

Most of the features utilized had a normal distribution. Figure VIII.4 and Figure VIII.5 in the Appendix depict the histograms of features and linear correlation between the daily maximum value and proposed measurements respectively. Most training features exhibited normal distributions and either positive or negative linearity, the representations of features noted are the strongest predictors for each model given by the Pearson Correlation Matrix in Figure VIII.9. The full training dataset on GitHub offers training data for the 36 monitors in the study region. Code is reproduceable for other study areas given proper API access to GEE and the EPA Air Quality Now website.

V.2.1. KNN Interpolation vs. Regression Based Imputation

Monitors which had less than 5% missing days over the initial timeframe were estimated. These missing values likely stem from either monitor malfunction or movement, many historical monitors were found to still be in use today (US EPA 2015). Interpolation and imputation strategies consisted of relating ground truth values from monitor measurements to their predicted value after replacement within the data set:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Interpolation Strategy** | **MSE** | **MAE** | **R2** | **RMSE** |
| Linear | 0.029 | 4.1 | 0.77 | 5.4 |
| ModAkima | 0.031 | 4.2 | 0.76 | 5.5 |
| Akima 1D | 0.031 | 4.3 | 0.75 | 5.5 |
| P(0) | 0.039 | 4.7 | 0.69 | 6.3 |
| P(2) | 2.4 | 48 | -18.46 | 49.4 |
| P(1) | 2.5 | 48 | -18.54 | 49.6 |
| P(2) | 2.6 | 48 | -19.93 | 51.3 |

Table V.2

Output of resulting estimations via KNN interpolation of the nearest 3 monitors. These errors could be drastically reduced via better spatial-temporal imputation methods as opposed to interpolation strategies. This is further discussed in Chapter VI. P(x) denotes polynomial interpolation of the x degree, strategies which were effective for variables like NDVI and Stratospheric NO2

The average deviation from each estimated value in the final dataset was around 4.1ppb ±5.4 ppb. The final distribution of all max values can be seen in the appendix under Figure VIII.TR.7, which splits the monitors into those which typically see lower than, at, and above average concentrations of max ozone concentrations detected with the AOI. Histograms for each of the training features can be in Figure VIII.TR.4 with the reduced version depicting theoretical training features in Figure 1. These were used in the final raster calculations due to their seasonal consistency and variance in sample ranges.

V.2. Individual Model Accuracy

Due to the stochasticity of surface ozone concentrations overtime (Figure 10), some seasons with low concentrations had minimal corrections made with residual kriging due to a near perfect prediction from the statistical ensemble. This is mainly due to the reaction’s reliance on incoming solar radiation and known seasonality of features as mentioned in Chapter 2. The error at surrounding points was better adjusted due to correlation with a simple, non-stochastic trend of geospatial uncertainty. The effects of residual kriging in this relatively small AOI still offered similar, albeit better representations of surface O3 reactions for dense urban areas such as those in (Al-Qassimi and Al-Salem 2020; Kong et al. 2023; Lo et al. 2024; Sadighi et al. 2018; Sun et al. 2022; Tong et al. 2017; Yin et al. 2019; Zou et al. 2019) and many more. Gradient boost yielded the best predictions suited for this project and depictions made in Chapter VI. Maps X and on represent the gradient boosted predictions and their RK enhanced versions to produce the represented surface. An excerpt from 2023 to 2024 is depicted to show a zoomed in variation of predicted values as compared to ground truths overtime:

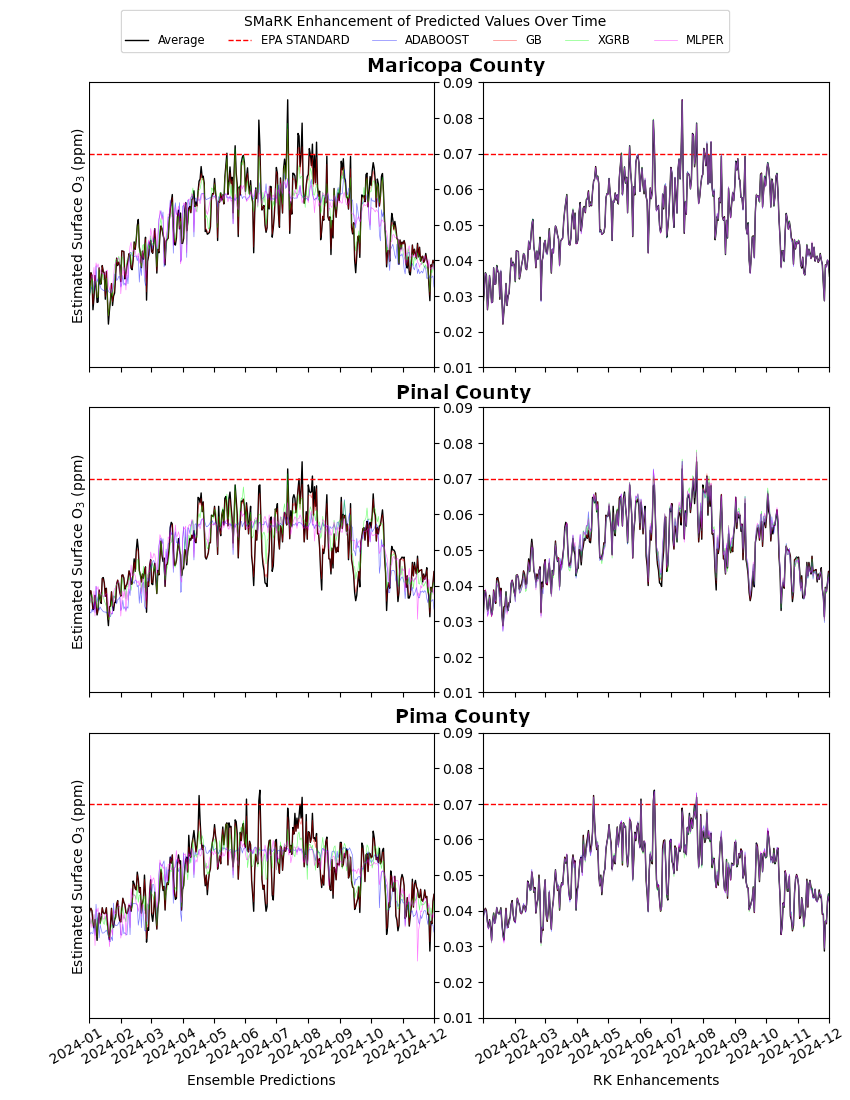
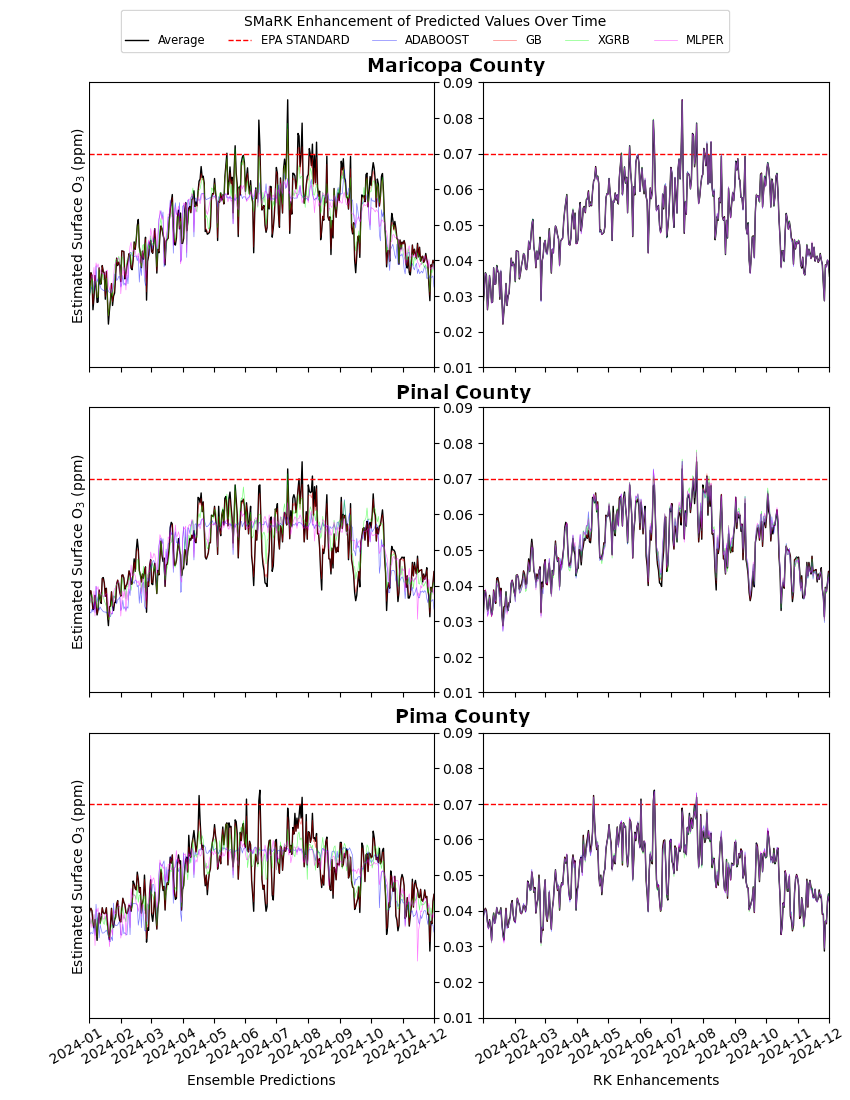
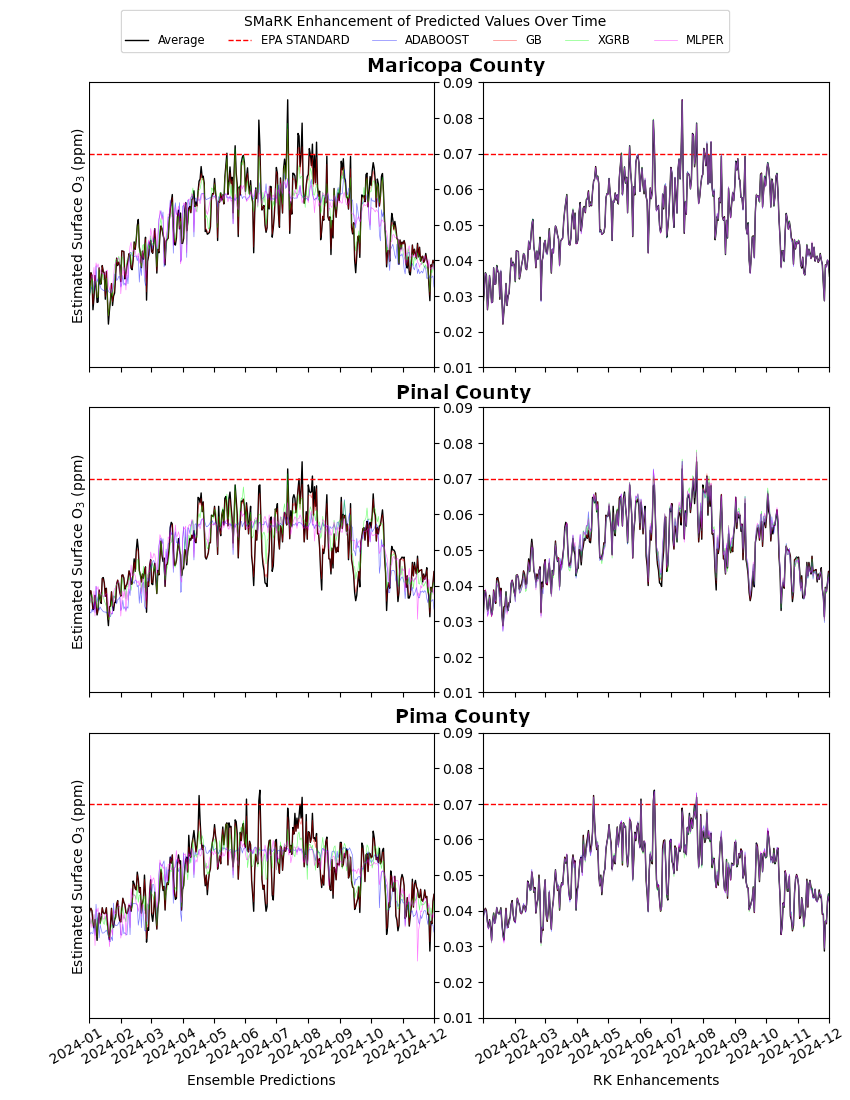
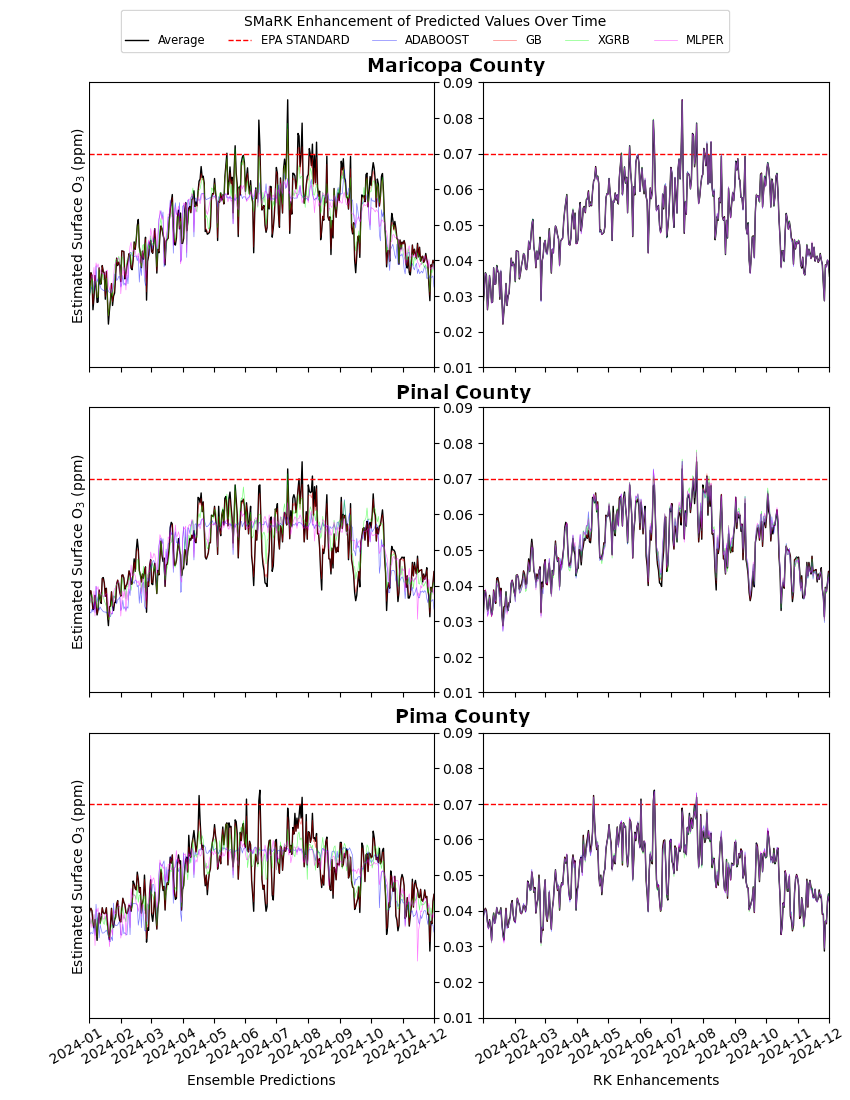


Figure V.2

Surface O3 approximations are overlaid with in-situ measurements overtime to depict differences in model error and improvements via RK. For each county, the left graph denotes statistical ensembles with no spatial interpolation of error covariates. The right plot denotes SMaRK methodology estimating these co-variates and incorporating them back into the overall approximation, better accounting for the lack of elevation changes in the AOI.

The full results over time are difficult to depict due to the resolution of the dataset as seen in Chapter VIII.2.9 and Chapter VIII.2.10. Overall, ensemble predictions alone required perfect parameters for effectiveness, else the learned trend was too general for accurate representations of fluctuations during seasons. Chapter VIII.2.10 depicts the variance among purely statistical methods for each county. Correlations for each feature set can be viewed Figure VIII.2.11-VIII.2.14 at better resolutions. When properly corrected for geospatial uncertainty with regression kriging, all methods were able to properly account for complex O3 trends. Ensemble methods alone seemed to fail during known low seasons of O3, where it shows high fluctuations with smaller correlations to lower variations with dependent variables.

Model tuning parameters, errors, and the final features selected for prediction are quickly mentioned, then further elaborated on in Chapter VI. ML/AI methods which performed the best were able to establish cyclic trends among the features across datasets. While some models themselves outperformed RK methods (e.g. random forest with the HD), the RK version of those methods performed the best out of the dataset. The full outputs for individual datasets and models can be seen in Tables VIII.5. and VIII.6 depicting statistical distributions for each. Excerpts from the figures created with these are

V.2.1. Adaptive Boost

Features tested with Adaptive boosting (ADA) offered the least predictive power with SMaRK error resulting in about 6.5% error and a mean average error of 6.0 ppb and MSE of 0.057 ppb. This was still better than negating spatial uncertainty completely, with 6.26% overall error as opposed to 12.48% overall error. The best dataset for ADA was the GOAT 25 features, likely due to the number of estimators set to 50. Unable to detect cyclic trends, this was best tuned to minimize linear loss with the associations found across surface O3 values in each sample and a constant learning rate of 1.0. Adaboost was the least frequently used method across literature in Chapter II, and this project showed as such. It’s likely that the simplistic linear associations could not be determined with many samples due to variations with surface O3 opposed to linearly co-funding features in G.O.A.T.24. Co-funding variables allowed for more deterministic variograms due to errors stemming from similar correlations as opposed to errors from positive and negatively correlated features. The least informative predictive dataset for ADA was MD with TD providing the most error with RK enhancements.

This dataset likely performed the best with G.O.A.T.24 due to the high usage of features for determining an overall linear loss trend estimation. RK was not as effective with ADA and TD since the features used were a combination of positive and negatively correlated features. TD didn’t show as strong of an initial complex trend, having the second most error with ADA alone. The best RMSE, MAE, MSE, MAPE, and R2 correlation coefficient for ADA compared to ADARK using G.O.A.T.24. was 4.80 ppb, 3.63 ppb, 0.0231 ppb, 8.07%, and 0.79 versus 3.38 ppb, 2.53 ppb, 0.012 ppb, 5.61%, and 0.898364 respectively. In addition, the overall maximum and minimum for ADA was furthest from actual predictions with a max and min of 87.41 ppb ± 4.80 ppb and 10.37 ppb ± 4.80 ppb respectively. The mean, standard deviation, and median were also more spread out with 48.26 ppb respectively. ADA from actual representations with a maximum of 89.58 ppb ± 3.38 ppb, minimum of 5.73 ppb ± 3.38 ppb, and mean value of 48.23 ppb ± 3.38 ppb.

V.2.2. Gradient Boost

The best application of Gradient Boosting (GB) yielded a tolerance of 0.0001, maximum number of 366 estimators with no limitation to sampling size. It was trained to minimize absolute error with a learning rate of 0.01 and no additional regularization weights. GB performed the best when trained to sample across yearly bins. This likely worked well due to proper incorporation of linear temporal aspects. Despite minimally increased overall percent accuracy by incorporating RK modelled uncertainty, the upper and lower estimates of the overall tendencies during high and low O3 seasons were estimated closer to the know trend at each monitor. In addition, the low usage of features from TD and trends across the AOI likely had greater trends across yearly samples. Increasing the parameters of the geo-field to include as many possible monitors would likely benefit the overall model, as consistent with (Goodchild 2011) and numerous other sources using point-sourced corrections for imagery e.g. (Adeeb and Shooter 2004; Cavaliere et al. 2023; Ionascu et al. 2021; Pires et al. 2008) to name a few. The best RMSE, MAE, MSE, MAPE, and R2 correlation coefficient for ADA compared to ADARK using G.O.A.T.24. was 4.80 ppb, 3.63 ppb, 0.0231 ppb, 8.07%, and 0.79 versus 3.38 ppb, 2.53 ppb, 0.012 ppb, 5.61%, and 0.898364 respectively. In addition, the overall maximum and minimum for ADA was 87.41 ppb ± 4.80 ppb and 10.37 ppb ± 4.80 ppb, with a mean of 48.26 ppb. XGBRK was closer to actual representations with a maximum of 89.58 ppb ± 3.38 ppb, minimum of 5.73 ppb ± 3.38 ppb, and mean value of 48.23 ppb ± 3.38 ppb.

V.2.3. Extreme Gradient Boost

Extreme Gradient Boosting (XGB) emerged as one of the most consistent and high-performing models across nearly all datasets. Its configuration favored the same high number of estimators as GB with 450, but a more moderate learning rate (~0.09). XGB effectively learned complex non-linear relationships in O₃ data and excelled particularly under TD and GOAT 25 features, achieving low overall error across each metric. Total errors ranging from 7.64% to 8.45% were reduced to 5.62% to 5.95% respectively. In addition, its adaptability to both theoretical simulations and high-feature datasets were better by itself, further corrected by RK error additions. The use of the ‘lossguide' grow policy allowed it to build deeper trees selectively, optimizing splits where the data was most informative. Overall, XGBoost proved especially valuable for applications where both feature richness and spatial-temporal variance were prominent like MD and TD. The best RMSE, MAE, MSE, MAPE, and R2 correlation coefficient for XBG compared to XGBRK using MD was 4.80 ppb, 3.63 ppb, 0.0231 ppb, 8.07%, and 0.79 versus 3.38 ppb, 2.53 ppb, 0.012 ppb, 5.61%, and 0.898364. The overall maximum and minimum for XGB was 87.41 ppb ± 4.80 ppb and 10.37 ppb ± 4.80 ppb, with a mean of 48.26 ppb. XGBRK was closer to actual representations with a maximum of 89.58 ppb ± 3.38 ppb, minimum of 5.73 ppb ± 3.38 ppb, and mean value of 48.23 ppb ± 3.38 ppb. MD was the best predictive dataset for XGBRK.

V.2.4. Random Forest

Random Forest consistently produced some of the lowest error rates, especially under the Historical and Theory datasets, where it achieved an MAE as low as 1.19 ppb and a minimal RMSE of 1.56 under the SMaRK methodology. Its effectiveness with relatively simple tuning — 70 trees and unrestricted depth — shows its innate strength as a bagging ensemble method that averages over diverse trees to reduce variance. The model excelled in datasets with more regular patterns and low noise, suggesting that it captures core linear and mildly non-linear patterns well. Importantly, Random Forest showed exceptional robustness under the SMaRK framework, maintaining stable performance even with spatial uncertainty and reduced training bias. Its ability to adapt without overfitting, even when regularization was absent, highlights its strength as a general-purpose model for environmental monitoring and point-source estimation tasks. This balance of accuracy and stability makes it a practical choice for operational forecasting systems. The best RMSE, MAE, MSE, MAPE, and R2 correlation coefficient for ADA compared to ADARK using G.O.A.T.24. was 4.80 ppb, 3.63 ppb, 0.0231 ppb, 8.07%, and 0.79 versus 3.38 ppb, 2.53 ppb, 0.012 ppb, 5.61%, and 0.898364 respectively. In addition, the overall maximum and minimum for ADA was 87.41 ppb ± 4.80 ppb and 10.37 ppb ± 4.80 ppb, with a mean of 48.26 ppb. XGBRK was closer to actual representations with a maximum of 89.58 ppb ± 3.38 ppb, minimum of 5.73 ppb ± 3.38 ppb, and mean value of 48.23 ppb ± 3.38 ppb.

V.2.5. Multi-Layered Perceptron

The Multi-Layered Perceptron (MLP) demonstrated moderate but inconsistent performance across datasets, showing potential in theoretical scenarios but struggling with real-world data, particularly in modern conditions. The deep architecture with four hidden layers (11–22–11–2) and adaptive learning via inverse scaling helped it learn non-linear trends effectively, but this came at the cost of generalizability. While the MLP achieved reasonably low errors in the Theory dataset (MAE ~2.5, RMSE ~3.34), it suffered from higher variability and larger total errors in modern datasets, likely due to sensitivity to noise, overfitting, or poor initialization. Its overall reduced error rate under the SMaRK method remained close to that of AdaBoost, indicating only marginal benefit from its complexity. The MLP could potentially be enhanced with more rigorous tuning, dropout, or batch normalization, but in this context, it lagged behind tree-based models in both accuracy and consistency. Thus, while theoretically powerful, MLP was less suitable for practical spatial-temporal O₃ estimation without further optimization. The best RMSE, MAE, MSE, MAPE, and R2 correlation coefficient for ADA compared to ADARK using G.O.A.T.24. was 4.80 ppb, 3.63 ppb, 0.0231 ppb, 8.07%, and 0.79 versus 3.38 ppb, 2.53 ppb, 0.012 ppb, 5.61%, and 0.898364 respectively. In addition, the overall maximum and minimum for ADA was 87.41 ppb ± 4.80 ppb and 10.37 ppb ± 4.80 ppb, with a mean of 48.26 ppb. XGBRK was closer to actual representations with a maximum of 89.58 ppb ± 3.38 ppb, minimum of 5.73 ppb ± 3.38 ppb, and mean value of 48.23 ppb ± 3.38 ppb.

V.3 Residual Kriging Enhancements

The residual kriging parameters utilized a Fourier series transformation to model potential data drift. Given the area has extremely low elevation change, drift was likely introduced due to the lack of latitudinal and longitudinal gradients. Since residual kriging accounts for spatial dependence of predicted errors, the relation of prediction, residual and location may depict variation over flat spaces given known trends like the location of monitors or it’s change in elevation (slope), causing spatial drift (Oliver and Webster 2014). Drift occurring from known spatial trends over the AOI can be utilized given the treatment of locations as auxiliary variables. such as in a GIS-based analysis of soil samples (A. Anand et al. 2021). One of the mathematical variations of kriging allowed to incorporate drift to better explain potential spatial dependance is known as Universal Kriging which has been long established (Chilès and Desassis 2018; Oliver and Webster 2014) and implemented in common geospatial libraries in R and Python (Michael Pyrcz 2024; Moraga 2024).

V.3.1. Drift Specifications

Initially, some machine learning models were able to adequately predict some seasons, with a very basic implementation of residual kriging. Upon further investigation and understanding of the model, all statistical models were vastly improved by RK of estimated error at each monitoring location. Gradient Boost and related models tend to underpredict values at monitoring locations. Extreme Gradient boost showed a tendency to predict repetitive values. All feature training sets were able to distinguish with enough variation to apply residual kriging, but more complex ensembles did not necessarily yield better results. The Random Forest and Multi-Layered Perceptron would tend to both over and under predict, adding more spatial correlation to the errors and improving accuracy when implemented back into the model by offering a more comprehensive trend of residual values.

The Fourier series implemented into the SMaRK method resulted in the best depiction of latitudinal and longitudinal drift when compared to point drift and specified drift in the kriging function. While elevation specified point drift and sinusoidal longitude and latitude drift still allowed for the model to account for large changes in surface ozone over small areas; the Fourier series can translate any wave equation into an infinite sum of sines and cosines. This allowed it to estimate the exact difference in surface O3 uncertainty at each pixel in the AOI given changes in latitude and longitude from monitor distributions. Functional drift with a Fourier series relationship of was ultimately the best at accounting for areas with either no monitors or large changes in detected surface O3 concentrations due to its ability to better correlate nearby monitor locations.

V.4. Computation Times

Overall, training the models on various datasets didn’t always improve model outcomes or computation times. Historical features were able to predict just as well as modern features when the RK drift function was properly tuned with a complex base. Table VIII.1.3 shows the full results of model training separated into each of the four datasets. Figure VIII.2 shows the results of RK enhanced predictions which better correlated with the overall trend. Gradient boost yielded the least error and ran second fastest next to extreme gradient boost (XRGB). XRGB consistently ran in a matter of seconds, insinuating that further tuning and testing is needed for GPU implementation into RF and MLPR methods. Large tree boosting methods and NNs Which took nearly 8 hours to 1 hour respectively. However, each of the other four models were able to be trained and validated with minimal effort and noting that patience is indeed a virtue.

The total computation time of data processing on the Mac was 24:15.47. Much of the script can be reduced to improve computations times on laptops and lower-power machines. This project was established with Python 3.9; keeping up with the numerous python versions was a lesson to the researcher in of itself. While keeping python up-to-date improving computation times from over 30 hours to a little over a day, the numerous other libraries which also required updating are likely suitable for python 3.14 and upcoming release of 3.15. The desktop in use did not implement the use of a GPU. While this would have dramatically reduced computation times, implementation of AMD’s ROCm software is currently in development, and implementation of this software is relatively new.

The total computation time of data processing with the Ryzen based processor was 14:15.47. Downloading, filtering, and processing of both vector and raster data took 00:46.11. Total computation time for model training, tuning, and prediction output was 01:37.32. Rk was completed with a computation time of 00:16.42. Image processing and final outputs took 01:57.52. The cloud-based desktop provided an NVIDIA based GPU, which allowed for Cuda implementation. Cuda is currently more effective in GPU AI/ML implementation than AMD’s ROCm software and utilizes more stable methods when computing with spatial data. The total computation time of data processing on the Intel x NVIDIA GPU combination was 14:15.47. Downloading, filtering, and processing of both vector and raster data took 00:46.11. Total computation time for model training, tuning, and prediction output was 01:37.32. RK was completed with a computation time of 00:16.42. Image processing and final outputs took 01:57.52.

V.5. Final Images and Outputs

The high-resolution surface O3 imagery captured urban spatial patterns as well as if not better than models of surface O3 using ML/AI techniques with CTM based features. Given the methodology is near the same, improvement can be associated with the incorporation of RK estimated values into the overall depiction of PHOTUC. While some models alone were able to predict with lower errors than with residual kriging implementations, the same statistical model with residual kriging improvements would have the lowest error overall. Random months to depict the full display of surface ozone maps for January 2019, March 2020, May 2021, July 2022, September 2023, and November 2024 can be viewed in Maps XI.3.7 and on from pages X to Y in the Appendix. All prior described trends, statistics, and overall results can be viewed prior to the maps in section XI.2. TD was the final dataset for use in PHOTUC, multivariate depictions of surface O3 and the first of each month with the associated financial year can be seen in Maps XI.3.4-XI.3.9; creating 5 unique maps with the best overall outcome from the SMaRK process.

V.5.1. GBRK Applications to PHOTUC

The best model as mentioned in V.2 was GBRK using TD features. Across all models, trends were better associated with the known value and overall distributions of Surface O3 in PHOTUC when accounting for spatial uncertainty. The TD results with GBRK predictions are depicted in Figure V.4. as an excerpt from Figure VIII.2.4. These features offered both positive and negative trends for the estimation of the field and allowed for the clearest separation of values in each sampling technique. In addition, model distributions per county were closer to the in-situ trend in both stand-alone and RK versions. Where the model was slightly overfitted, RK estimations of residuals were able to account for the variations in urban locations and high elevations spikes in the AOI.

A screenshot of a graph

AI-generated content may be incorrect.Further work and incorporation of this method into spatial programs is necessary to unleash the full capacity of GIS into the modern world, as these errors were lower than that of similar CTM and Emissions based representations of surface O3 concentrations e.g.(Chattopadhyay, Midya, and Chattopadhyay 2019; Pan, Harrou, and Sun 2023; Zhang et al. 2022). This project yielded low-errors even with comparatively lack-luster incorporation of training values representing surface O3. Proper implementation of GPU workloads and newly updated python packages would allow for stronger

This thesis has undergone 3 iterations of python updates, with its final release on GitHub having stability in python 3.12.10. Python 3.13 was released during this work on Oct. 7, 2024, and was not properly implemented due to this string of upgrades to the language that were not yet implemented into libraries like pykrige. Other languages such as R and JavaScript also offer similar utility, perhaps offering more suitable processing times on laptops and rapid processing times on high-end desktop computers. The resulting script in python produced TIF files which were imported into ArcGIS Pro to demonstrate proficiency with the program and establish the precursors to full implementation of the script for public and private entities using ESRI’s mapping suite.

V.5.2. County Demographics of PHOTUC

Further breaking the results down by county statistics allows for insights into who is affected and by how much surface O3 as estimated by SMaRK. In addition, combining the AOI with its distributions of household and income portrays numerous trends mentioned in Chapter II. Figure V.1 displays the first day of thepredicted dataset from GBRK methods. The rest of the daily images can be seen in the Appendix. Individual maps of population distributions, income, and surface O3 in the AOI can further viewed in Map XI.3.1, Map XI.3.2, and Maps XI.3.7 on. Population data for the year 2020 was used and not updated for this project, as yearly population updates are held for future work.

Yearly income data from 2019 to 2023 depict estimated financial changes over the time frame. All daily surface O3 rasters for each month can be seen starting on Map XI.3.13 and on.The next chapter covers a detailed description of surface O3 trends within the urban centers of Phoenix and Tucson in Arizona to show the effectiveness of these results. Maps of population and income are depicted with bivariate color schemes in which the combination of two colors represents the values of two different fields. For income representations, the deviation between the median and mean income by census tract is represented as yellow. The total percentage of occupied households is depicted as the estimated total amount of occupied households divided by the total number of available homes in each tract and colored blue. Population density and distributions were colored in a similar manner, replacing income deviations with population per 100k; blue will represent the same household percentage used in the population distribution maps. When combined, these two colors merge into a green hue wherein the county depicts high deviations of mean income to median income and a high population per capita.

The textures are left in greyscale for the initial portrayal of these maps; they are later overlaid with a continuous red color scheme of surface O3 values, shifting yellow, green, and blue to shades of orange, yellow, and purple respectively. While the core statistics behind the coloring scheme may seem a little complex, they are not without benefits, depicting interesting spatial trends within PHOTUC. In essence, three main colors are noted as key points within each map. Blue shaded census tracts and census tract groups are representations of communities which have a high percentage of occupied homes. Overall, 16 maps were created with fine detail to show the potential of this work. When the bivariate color schemes denoting census data are properly overlaid with

V.5.3. Omission of Certain Methods

Given the scale of the final predictive value, Box-Cox or Yeo-Johnson values were not utilized. Upon the initial drafting of this thesis, normalized values yielded minimal improvements in the overall accuracy of each model and confused each model by keeping each value within relatively the same range. Scaling the dataset offered the same deviations as non-scaled data without having to implement another step in an otherwise overall arduous scheme. In addition, the utility offered in the python functions allow for daily predictions at larger AOIs. Development of larger tiles, inclusion of more monitoring systems, and historical trend depictions are saved for future work by the researcher.

Citations

Adeeb, F, and D Shooter. 2004. “Variation of Surface Ozone in the Ambient Air of Auckland, New Zealand.” *ENVIRONMENTAL MONITORING AND ASSESSMENT* 95 (1–3): 201–20. doi:10.1023/B:EMAS.0000029904.28706.c0.

Al-Qassimi, M, and SM Al-Salem. 2020. “Ozone (O3) Ambient Levels as a Secondary Airborne Precursor in Fahaheel Urban Area, the State of Kuwait.” *ATMOSPHERIC SCIENCE LETTERS* 21 (9). doi:10.1002/asl.983.

Anand, Akash, Prachi Singh, Prashant K. Srivastava, and Manika Gupta. 2021. “GIS-Based Analysis for Soil Moisture Estimation via Kriging with External Drift.” In *Agricultural Water Management*, 391–408. Elsevier. doi:10.1016/b978-0-12-812362-1.00019-9.

Anand, KAA, D Ganguly, S Ghosh, VK Soni, and S Dey. 2025. “Insights into Surface Ozone Variability in India (1980-2014) through CMIP6 Model Analysis.” *ATMOSPHERIC ENVIRONMENT* 351 (June). doi:10.1016/j.atmosenv.2025.121180.

C T, Resmi, Nishanth T, Satheesh Kumar M K, Balachandramohan M, and Valsaraj K T. 2020. “Long-Term Variations of Air Quality Influenced by Surface Ozone in a Coastal Site in India: Association with Synoptic Meteorological Conditions with Model Simulations.” *Atmosphere* 11 (2). MDPI: 193–193. doi:10.3390/atmos11020193.

Cavaliere, A, L Brilli, BP Andreini, F Carotenuto, B Gioli, T Giordano, M Stefanelli, C Vagnoli, A Zaldei, and G Gualtieri. 2023. “Development of Low-Cost Air Quality Stations for next-Generation Monitoring Networks: Calibration and Validation of NO2 and O3 Sensors.” *ATMOSPHERIC MEASUREMENT TECHNIQUES* 16 (20): 4723–40. doi:10.5194/amt-16-4723-2023.

Chattopadhyay, Goutami, Subrata Kumar Midya, and Surajit Chattopadhyay. 2019. “MLP Based Predictive Model for Surface Ozone Concentration over an Urban Area in the Gangetic West Bengal during Pre-Monsoon Season.” *Journal of Atmospheric and Solar-Terrestrial Physics* 184 (March). Elsevier BV: 57–62. doi:10.1016/j.jastp.2019.01.008.

Chen, B, YX Wang, JP Huang, L Zhao, RM Chen, ZH Song, and JS Hu. 2023. “Estimation of Near-Surface Ozone Concentration and Analysis of Main Weather Situation in China Based on Machine Learning Model and Himawari-8 TOAR Data.” *SCIENCE OF THE TOTAL ENVIRONMENT* 864 (March). doi:10.1016/j.scitotenv.2022.160928.

Chilès, Jean-Paul, and Nicolas Desassis. 2018. “Fifty Years of Kriging.” In *Handbook of Mathematical Geosciences*, 589–612. Cham: Springer International Publishing. doi:10.1007/978-3-319-78999-6\_29.

Gillies, Sean and others. 2013. “Rasterio: Geospatial Raster I/O for Python Programmers.” Mapbox. https://github.com/rasterio/rasterio.

Goodchild, Michael F. 2011. “Scale in GIS: An Overview.” *Geomorphology* 130 (1–2). Elsevier BV: 5–9. doi:10.1016/j.geomorph.2010.10.004.

Hu, XM, J Zhang, WH Xue, LH Zhou, YF Che, and T Han. 2022. “Estimation of the Near-Surface Ozone Concentration with Full Spatiotemporal Coverage across the Beijing-Tianjin-Hebei Region Based on Extreme Gradient Boosting Combined with a WRF-Chem Model.” *ATMOSPHERE* 13 (4). doi:10.3390/atmos13040632.

Huang, HY, RK Zou, DY Li, LG Zhou, C Ren, FM Zeng, and XY Dai. 2025. “Estimation of Full-Coverage Surface Ozone Using Sentinel-5P Data and a GAN-LGBM Model in the YRDUA, China.” *REMOTE SENSING LETTERS* 16 (2): 136–45. doi:10.1080/2150704X.2024.2440667.

Ionascu, ME, N Castell, O Boncalo, P Schneider, M Darie, and M Marcu. 2021. “Calibration of CO, NO2, and O3 Using Airify: A Low-Cost Sensor Cluster for Air Quality Monitoring.” *SENSORS* 21 (23). doi:10.3390/s21237977.

Kleinert, F, LH Leufen, and MG Schultz. 2021. “IntelliO3-Ts v1.0: A Neural Network Approach to Predict near-Surface Ozone Concentrations in Germany.” *GEOSCIENTIFIC MODEL DEVELOPMENT* 14 (1): 1–25. doi:10.5194/gmd-14-1-2021.

Ko, K, S Cho, and RR Rao. 2022. “Machine-Learning-Based Near-Surface Ozone Forecasting Model with Planetary Boundary Layer Information.” *SENSORS* 22 (20). doi:10.3390/s22207864.

Kong, SJ, T Wang, F Li, JJ Yan, and ZG Qu. 2023. “Unraveling Spatiotemporal Patterns and Multiple Driving Factors of Surface Ozone across China and Its Urban Agglomerations Management Strategies.” *FRONTIERS IN ECOLOGY AND EVOLUTION* 11 (June). doi:10.3389/fevo.2023.1103503.

Li, MY, SC Yu, X Chen, Z Li, YB Zhang, LQ Wang, WP Liu, et al. 2021. “Large Scale Control of Surface Ozone by Relative Humidity Observed during Warm Seasons in China.” *ENVIRONMENTAL CHEMISTRY LETTERS* 19 (6): 3981–89. doi:10.1007/s10311-021-01265-0.

Li, QY, J Li, ZX Wang, B Liu, W Wang, and ZF Wang. 2024. “Development of a City-Level Surface Ozone Forecasting System Using Deep Learning Techniques and Air Quality Model: Application in Eastern China.” *ATMOSPHERIC ENVIRONMENT* 339 (December). doi:10.1016/j.atmosenv.2024.120865.

Lo, KC, WH Cheng, C Lin, CH Hung, CS Yuan, and YL Tseng. 2024. “Elevated Surface Ozone Concentration Caused by Subtropical Cyclones and Topographical Effect: Model Simulation and Field Measurement.” *URBAN CLIMATE* 57 (September). doi:10.1016/j.uclim.2024.102093.

Michael Pyrcz. 2024. “GeostatsGuy/GeostatsPyDemos\_Book: V0.2.” Zenodo. doi:10.5281/ZENODO.15169133.

Moraga, Paula. 2024. *Spatial Statistics for Data Science: Theory and Practice with R*. First edition. Chapman & Hall/CRC Data Science Series. Boca Raton: CRC Press, Taylor & Francis Group.

Mu, X, SC Wang, P Jiang, and YL Wu. 2023. “Estimation of Surface Ozone Concentration over Jiangsu Province Using a High-Performance Deep Learning Model.” *JOURNAL OF ENVIRONMENTAL SCIENCES* 132 (October): 122–33. doi:10.1016/j.jes.2022.09.032.

Nelson, D, Y Choi, B Sadeghi, AK Yeganeh, M Ghahremanloo, and J Park. 2023. “A Comprehensive Approach Combining Positive Matrix Factorization Modeling, Meteorology, and Machine Learning for Source Apportionment of Surface Ozone Precursors: Underlying Factors Contributing to Ozone Formation in Houston, Texas.” *ENVIRONMENTAL POLLUTION* 334 (October). doi:10.1016/j.envpol.2023.122223.

Oliver, M.A., and R. Webster. 2014. “A Tutorial Guide to Geostatistics: Computing and Modelling Variograms and Kriging.” *CATENA* 113 (February). Elsevier BV: 56–69. doi:10.1016/j.catena.2013.09.006.

Pan, Qilong, Fouzi Harrou, and Ying Sun. 2023. “A Comparison of Machine Learning Methods for Ozone Pollution Prediction.” *Journal of Big Data* 10 (1). Springer Science and Business Media LLC. doi:10.1186/s40537-023-00748-x.

Pires, JCM, SIV Sousa, MC Pereira, MCM Alvim-Ferraz, and FG Martins. 2008. “Management of Air Quality Monitoring Using Principal Component and Cluster Analysis -: Part II:: CO, NO2 and O3.” *ATMOSPHERIC ENVIRONMENT* 42 (6): 1261–74. doi:10.1016/j.atmosenv.2007.10.041.

Rossum, Guido van, and Fred L. Drake. 2010. *The Python Language Reference*. Release 3.0.1 [Repr.]. Python Documentation Manual / Guido van Rossum; Fred L. Drake [Ed.], Pt. 2. Hampton, NH: Python Software Foundation.

Rouault, Even, Frank Warmerdam, Kurt Schwehr, Andrey Kiselev, Howard Butler, Mateusz Łoskot, Tamas Szekeres, et al. 2025. “GDAL.” Zenodo. doi:10.5281/ZENODO.15375292.

Sadighi, K, E Coffey, A Polidori, B Feenstra, Q Lv, DK Henze, and M Hannigan. 2018. “Intra-Urban Spatial Variability of Surface Ozone in Riverside, CA: Viability and Validation of Low-Cost Sensors.” *ATMOSPHERIC MEASUREMENT TECHNIQUES* 11 (3): 1777–92. doi:10.5194/amt-11-1777-2018.

Seroji, Abdulaziz R. 2016. “The Ground Ozone Variations with UV Radiation during Winter and Spring Seasons in 2007 over Makkah.” *Journal of King Abdulaziz University: Science* 29 (1). King Abdulaziz University, Scientific Publishing Centre: 35–55. doi:10.4197/Sci.22-1.3.

Sun, HT, YM Shin, MT Xia, SX Ke, M Wan, L Yuan, YM Guo, and AT Archibald. 2022. “Spatial Resolved Surface Ozone with Urban and Rural Differentiation during 1990-2019: A Space-Time Bayesian Neural Network Downscaler.” *ENVIRONMENTAL SCIENCE & TECHNOLOGY* 56 (11): 7337–49. doi:10.1021/acs.est.1c04797.

Tong, L, HL Zhang, J Yu, MM He, NB Xu, JJ Zhang, FZ Qian, JY Feng, and H Xiao. 2017. “Characteristics of Surface Ozone and Nitrogen Oxides at Urban, Suburban and Rural Sites in Ningbo, China.” *ATMOSPHERIC RESEARCH* 187 (May): 57–68. doi:10.1016/j.atmosres.2016.12.006.

US EPA. 2015. *The Benefits and Costs of the Clean Air Act from 1990 to 2020, Final Report, Revision A, April 2011*. The Benefits and Costs of the Clean Air Act from 1990 to 2020. U.S. Environmental Protection Agency Office of Air and Radiation.

Yin, CQ, XJ Deng, Y Zou, F Solmon, F Li, and T Deng. 2019. “Trend Analysis of Surface Ozone at Suburban Guangzhou, China.” *SCIENCE OF THE TOTAL ENVIRONMENT* 695 (December). doi:10.1016/j.scitotenv.2019.133880.

Zhang, Wenxiu, Di Liu, Hanqin Tian, Naiqin Pan, Ruqi Yang, Wenhan Tang, Jia Yang, et al. 2022. “Recurrent Mapping of Hourly Surface Ozone Data (HrSOD) across China during 2005-2020 for Ecosystem and Human Health Risk Assessment.” *Earth System Science Data Discussions*, December. Copernicus Gesellschaft mbH, 2–36. doi:10.5194/essd-2022-428.

Zhao, H, YF Zheng, and C Li. 2018. “Spatiotemporal Distribution of PM2.5 and O3 and Their Interaction During the Summer and Winter Seasons in Beijing, China.” *SUSTAINABILITY* 10 (12). doi:10.3390/su10124519.

Zou, Y, XJ Deng, T Deng, CQ Yin, and F Li. 2019. “One-Year Characterization and Reactivity of Isoprene and Its Impact on Surface Ozone Formation at A Suburban Site in Guangzhou, China.” *ATMOSPHERE* 10 (4). doi:10.3390/atmos10040201.